Deep Learning-based Land Use and Land Cover Changes Detection from Satellite Imagery : a case study of the city of Richard Toll

Mandicou BA
Cheikh Anta Diop University (UCAD), Polytechnic Higher School (ESP), Research Institute for Development (IRD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
mandicou.ba@esp.sn

Pape Ibrahima THIAM
Cheikh Anta Diop University (UCAD), Polytechnic Higher School (ESP), Research Institute for Development (IRD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
papeibrahimathiam@esp.sn

Etienne DELAY
French Agricultural Research Centre for International Development (CIRAD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
etienne.delay@cirad.fr

Charles Abdoulaye NGOM
Cheikh Anta Diop University (UCAD), Polytechnic Higher School (ESP), Research Institute for Development (IRD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
charlesabdoulayengom@esp.sn

Idy DIOP
Cheikh Anta Diop University (UCAD), Polytechnic Higher School (ESP), Research Institute for Development (IRD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
idy.diop@esp.sn

Alassane BAH
Cheikh Anta Diop University (UCAD), Polytechnic Higher School (ESP), Research Institute for Development (IRD), Unit of Mathematical Modeling and Computer Science of Complex Systems (UMI UMMISCO - UCAD)
Dakar, SENEGAL
alassane.bah@ucad.edu.sn

ABSTRACT

In this paper, we propose the detection of land use and land cover changes from satellite imagery taken in Richard Toll. The Senegal River Valley, particularly the region encompassing Richard Toll, presents a significant research interest due to the prevalence of extensive agro-industrial activities. These activities induce profound alterations in the vegetative landscape, particularly evident upon their initiation or during expansion phases. Concurrently, these regions are obligated to reconcile the exigencies of pastoral sustainability. The identification of land use modifications through change detection in these areas is crucial for the prognostication and management of potential socio-environmental conflicts. Our approach is based on Deep Learning models applied to the analysis of satellite images, falling within the field of remote sensing where we automate the process of satellite images segmentation before tackling the generation of changes map. The methodology begins with the collection of geospatial-temporal data, 3-channel images taken at different points in time and in different spaces, of the area of interest via Google Earth Pro. The study region is divided into eight distinct classes, including cultivated fields, uncultivated fields, land, water, buildings, roads, football fields and vegetation. *U-Net* and *FCN-8* deep learning architectures are used to achieve that goal by generating the segmented masks in order to highlight the changes areas by creating changes map during a post-process. We compare these two models and opt for the *U-Net* model, which offers the best performances.

KEYWORDS

Deep Learning, U-Net, FCN-8, Remote sensing, monitoring, Change Detection, Richard Toll, Senegal.

ACM Reference Format:
Mandicou BA, Pape Ibrahima THIAM, Etienne DELAY, Charles Abdoulaye NGOM, Idy DIOP, and Alassane BAH. 2024. Deep Learning-based Land Use and Land Cover Changes Detection from Satellite Imagery : a case study of the city of Richard Toll. In *2024 The 7th International Conference on Machine Vision and Applications (ICMVA 2024), March 12–14, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3653946.3653956

1 INTRODUCTION

This study delves into the pivotal role of change detection in the realm of satellite imagery analysis, particularly in the context of mitigating and anticipating social tensions between agricultural and pastoral communities. The crux of our research is centered on identifying shifts in land use patterns, which are often precursors to conflict between farmers and herders over resource allocation. By pinpointing these changes, especially in regions where agricultural expansion or intensification intersects with traditional pastoral routes, we aim to provide early warnings and facilitate timely interventions. However, detecting these changes is just the initial step.

ICMVA 2024, March 12–14, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-1655-3/24/03
https://doi.org/10.1145/3653946.3653956

It's imperative to also consider the diverse forms of support and assistance required by the impacted populations. This entails devising strategies not only to mitigate immediate disputes but also to ensure sustainable coexistence and resource sharing. Through this approach, we aspire to contribute towards a more harmonious balance between agricultural development and pastoral needs, thereby fostering socio-environmental resilience in these sensitive regions.

The comparison between traditional pixel-based classification methods [6, 17, 18, 20] and newer object-based methods for Land Use/Land Cover (LULC) classification were sometime used [9, 12, 13, 15, 29]. However, the shift towards Deep Learning is preferred for its ability to maintain semantic information and automate the process [5, 14, 22, 23]. The adopted method involves the application of deep learning techniques for classifying geospatial-temporal satellite images within the field of remote sensing.

Therefore, our study is based on the innovative approaches includes careful geospatial-temporal data collection in the Richard Toll region using Google Earth Pro software. The areas of interest is divided into eight distinct classes: cultivated fields, uncultivated fields, land, water, buildings, roads, football fields, and vegetation, aligning with the GGW objectives for promoting sustainable land use and land cover.

The remainder of this paper unfolds as follows. Section 2 undertakes an exploration of related works within the field, offering a comprehensive contextual backdrop for our research. Section 3 meticulously elucidates the methodologies that constitute the data acquisition strategy step, the deep learning methods adopted, fine-tuning strategy and process to detect change land cover changes. Section 4 unveil the results derived from the outlined methodologies, coupled with an exhaustive analysis and subsequent discussion in Section 5, in order to provide a nuanced understanding of the implications and contributions of our research. Finally, we conclude and present some research perspectives in Section 6.

2 RELATED WORKS

In this section, we examine existing literature related to the detection of land use and land cover changes using deep learning techniques. Increasing attention is being directed towards the utilization of deep learning techniques for generating land cover change maps. A substantial body of research is dedicated to exploring the application of deep learning in this context.

In [11], Ebel et *al.* have developed a multimodal Siamese architecture for change detection. The network consists of two encoder branches for each sensor and a decoder part integrating the functionalities of the previous layers. Each encoder branch processes the bitemporal signal of its corresponding modality. samples, SAR and multispectral optics, in two passes. The extracted features are transmitted over skip connections in a U-Net like mode then concatenated.

A high-performing change detection algorithm using a U-Net architecture on High-Resolution Multi-Spectral (HRMS) images has been proposed in [33]. They enhanced the algorithm with a low-complexity pan-sharpening method, enabling joint use of panchromatic and multi-spectral images. The feature-level U-Net model included a subtraction network for extracting dynamic difference (DI) images for low-level and high-level features. Results demonstrated superior performance with an average F-1 score of 0.62, a percent correct classification (PCC) of 98.78%, and a kappa coefficient of 61.60% for test datasets. However, potential limitations, like dependence on HRMS data and variations in image quality, may affect performance in different scenarios.

In [30], the author investigates the use of Convolutional Neural Networks (CNN) for land cover classification from very high-resolution multispectral orthorectified images. Leveraging the abundance of Very High Resolution (VHR) data, the study addresses challenges related to varying appearances of land features due to factors like shooting time, sensor settings, image processing, and geographic context. The importance of testing classifier generalizability on unseen datasets is emphasized. Results indicate promising use of CNN for land cover classification, achieving an average accuracy rate of 96% on the UC Merced dataset with 5-fold cross-validation. The model's generalization capability is demonstrated on unrelated VHR images beyond the initial training data. Addressing limitations, the study acknowledges the challenge of obtaining large labeled datasets for deep learning in land cover classification. Techniques like "refinement learning" for reusing trained networks on other tasks are explored. Variability in land feature appearances is recognized as another complexity in classification [30].

In [8], Sentinel-2 time series data from the European Space Agency (ESA) were utilized for land cover classification. The data, covering the 2017-2018 agronomic year, included multispectral images in 13 bands with varying spatial resolutions. Fourteen features were extracted from 30 time steps, and field data from the Ministry of Agriculture were used for model training and testing. A Bidirectional Long-Short-Term Memory (BiLSTM) network with two layers and a softmax output layer was employed. The 2-BiLSTM model's mask exhibited a spatial distribution of land cover classes aligning with Spain's Parcel Identification System (LPIS). Identified categories included fruit trees, arable land, rice crops, and natural vegetation. The model achieved an overall accuracy of 98.7% against the test set, outperforming other segmentation models. Precision, recall, and F-1 scores varied for each class, with notable performance on natural vegetation but slightly lower discriminatory power for permanent crops compared to annual crops [8].

In [4], researchers proposed a non-overlapping grid-based approach to train FCN-8 with VGG-16 weights for segmenting satellite images into forest, built-up areas, agricultural land, and water classes. FCN-8 utilizes discriminative features learned by the lower resolution encoder, projecting them into higher resolution pixel space for dense classification. The Gaofen-2 dataset, comprising 150 images from over 60 Chinese cities, was used for experimentation. The proposed approach significantly outperformed eCognition software, achieving an average accuracy of 91.0% and an average intersection over union (IoU) of 0.84. In contrast, eCognition software attained an average accuracy of 74.0% and an average IoU of 0.60. While FCN-8 demonstrated superior performance, limitations included boundary errors between land cover classes, analyzed in detail. Model performance relies on satellite image quality and resolution, with lower quality images potentially affecting accuracy. Geographic diversity in the studied regions can also impact model performance [4].

Figure 1: Some satellite images collected from the Richard Toll region

3 METHODS

In this section, we outline our study's methodology, covering the acquisition, annotation, augmentation, and preprocessing of three-channel satellite images. We discuss model selection for segmentation of images where a label is given to each pixel of an image [16], fine-tuning [21], and creation of change maps. Strategies for addressing data limitations and imbalance are detailed. A comparative analysis of chosen models provides insights into their precision.

3.1 Dataset Acquisition

In building our dataset for the Great Green Wall initiative, we strategically focused on gathering satellite images from Richard Toll in Senegal, a key agricultural hub known for vast sugar cane fields. This region's significance in agriculture makes it ideal for studying the initiative's impact.

The collected satellite imagery, obtained at various resolutions using Google Earth Pro, offers valuable insights into land use and vegetation patterns (see Fig 1). This diverse dataset enhances the versatility of our research for robust analysis in subsequent stages.

3.2 Dataset Annotation

In the data preparation process, after initial image collection, our crucial step involves annotating images using Labelme, a versatile tool [19]. Labelme simplifies the process, allowing us to draw polygons, assign labels, and capture information.

(a) Colors of classes **(b) Satellite images annotation**

Figure 2: Illustration of the annotation process

This annotated dataset, as illustrated in Fig. 2b, is essential for training and evaluating our deep learning algorithms. A Python script, coupled with Labelme's JSON file, aids in creating masks representing the labels.

3.3 Dataset Augmentation

To diversify our dataset, TensorFlow's `ImageDataGenerator` applied controlled transformations like rotation, shifts, shearing, zooming, and flips. Integrated seamlessly into our workflow, this tool efficiently expands the dataset, enhancing our deep learning model's effectiveness. It overcomes small dataset limitations, promoting generalization to unseen data—an essential aspect for our research success.

3.4 Dataset Preprocessing

In preprocessing step, we address image size and shape with resizing and cropping for uniformity. Images are read, converted to RGB, and resized. The « patchify function » divides resized images into patches for further analysis. Similarly, mask images are resized appropriately. Processed image and mask patches are collected for subsequent research. Data scaling normalizes pixel values for consistency and aids model convergence during training. This preprocessing pipeline ensures dataset formatting and standardization [2]. To do this, we evaluate the normalized value $\mathcal{X}_{\text{normalized}}$ of pixel image following by Eq 1.

$$\mathcal{X}_{\text{normalized}} = \frac{\mathcal{X} - \mathcal{X}_{\text{min}}}{\mathcal{X}_{\text{max}} - \mathcal{X}_{\text{min}}} \tag{1}$$

Here, $\mathcal{X}$ is the all set of image pixel, $\mathcal{X}_{\text{normalized}}$ is the normalized value of $\mathcal{X}$, and $\mathcal{X}_{\text{min}}$ and $\mathcal{X}_{\text{max}}$ are the minimum and maximum values in the set of image pixel.

3.5 Deep Learning Adopted Models

For our study, we have opted for two state-of-the-art Convolutional Neural Network (CNN) architectures tailored for semantic segmentation tasks:

3.5.1 ***U-Net with Inception ResNetV2 Encoder.*** Our initial model draws inspiration from the highly acclaimed U-Net architecture, a stalwart in the field of semantic segmentation [28]. This model, as shown in Fig. 3, leverages the formidable Inception ResNetV2 as its encoder, creating a potent combination of two state-of-the-art architectures. U-Net is widely acknowledged for its proficiency in capturing intricate spatial features, particularly well-suited for tasks that demand precise delineation of object boundaries.

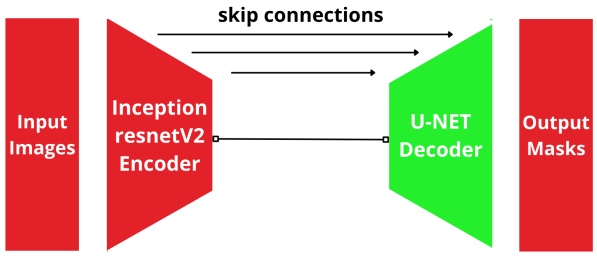


Figure 3: U-Net architecture with Inception resnetv2 as encoder

The choice of Inception ResNetV2 as the encoder brings additional strength to our model. Inception ResNetV2 is renowned for its capability to extract hierarchical features, allowing it to discern both local details and global contextual information. This synergy between U-Net and Inception ResNetV2 results in a robust architecture that excels in understanding the complex relationships within an image.

By seamlessly integrating the strengths of U-Net and Inception ResNetV2, our model is poised to tackle segmentation tasks with a heightened level of accuracy and efficiency. The Fig 3 illustrates the overall architecture, showcasing the effective collaboration between U-Net's specialized segmentation capabilities and the feature-rich hierarchical extraction prowess of Inception ResNetV2

3.5.2 ***FCN-8 with VGG19 Encoder.*** Our second model embraces the Full Convolutional Network (FCN-8) [4], a powerful architecture designed for end-to-end pixel-wise predictions. Enhancing its capabilities, we've equipped FCN-8 with a VGG19 encoder, a proven and effective feature extractor that complements the architecture seamlessly (refer to Fig 4). VGG19, known for its simplicity and efficiency, lays a solid foundation for robust feature extraction, setting the stage for precise segmentation tasks.

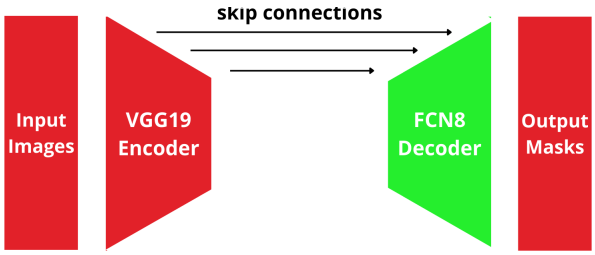


Figure 4: FCN-8 architecture with VGG19 as encoder

The '8' in FCN-8 means the upsampling factor, a crucial aspect that empowers the network to generate finely detailed segmentation maps. This architecture excels in capturing intricate information spanning the entirety of the input image. Its ability to maintain spatial details makes FCN-8 particularly well-suited for tasks that demand nuanced pixel-wise predictions.

In selecting these models, we strike a balance between advanced feature extraction and efficient segmentation. The distinctive architectures of U-Net and FCN-8 make them invaluable candidates for our research study. By incorporating these models, we aim to delve into a comprehensive exploration and comparison of their performance within the unique context of our dataset and research objectives.

3.6 Training

In this section, we present how our adopted approaches are trained and different parameters that were used in order to do this.

Firstly, we randomly initialized the parameters of U-Net [28] and FCN-8 [4]. We proceeded to train them by using the Adam optimizer that is a widely used optimizer in deep learning, combining the benefits of *SGD* and *RMSprop* by automatically adjusting the learning rate for each parameter, enabling fast convergence and better adaptation to complex loss landscapes [3]. The learning rate is fixed to 0.001 for the beginning. The activation function used in all the layers of both models was *Rectified Linear Unit* (*ReLU*) (see Eq 3) except the outputs layers of both where a $Softmax$ function (see Eq 2) was used to get a probability distribution of output classes [7, 10, 27].

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{2}$$

$$ReLU(x) = \max(0, x) \tag{3}$$

As we have unbalanced classes, we choose $Dice_{Loss}$ (see Eq 4) and $Focal_{Loss}$ (see Eq 5) and we sum them in order to use their total (see Eq 6) as loss function ($Total_{Loss}$) to overcome the unbalanced problems [31].

$$Dice_{Loss} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4}$$

Where TP represents the True Positives, FP represents the False Positives FN represents the False negatives [31].

$$Focal_{Loss} = -(1 - \hat{y})^{\gamma} \cdot \log(\hat{y}) \tag{5}$$

Where $\hat{y}$ is the probability predicted by the model for the target class, γ is a modulation parameter that can be adjusted [31].

$$Total_{Loss} = Dice_{Loss} + Focal_{Loss} \tag{6}$$

3.7 Fine-Tuning Strategy

Following the initial training phase, we employ a fine-tuning strategy to enhance the performance and adapt the models to the specifics of our dataset. The fine-tuning process involves a targeted adjustment of the model's weights, focusing on layers close to the input and output [21].

For both the U-Net with Inception ResNetV2 [28] and FCN-8 with VGG19 [4] models, we adopt a two-pronged approach:

3.7.1 Blocking Layers Close to the Input. In the initial stages of fine-tuning, we freeze or « block » layers close to the input. By preserving the knowledge encoded in these early layers, we ensure that the models retain their ability to extract fundamental features from the input data. This step is crucial for maintaining the models' understanding of low-level patterns and structures present in the images.

3.7.2 Updating Layers Close to the Output. Concurrently, we selectively update and fine-tune layers closer to the output. This targeted adjustment allows the models to adapt to the specific characteristics of our dataset, learning high-level representations that are more tailored to the nuances present in our segmentation task. By updating these layers, the models can refine their predictions and optimize performance for our specific context.

The fine-tuning process strikes a balance between leveraging pre-trained knowledge and adapting to the intricacies of our data. This iterative approach contributes to the models' ability to generalize well and achieve superior segmentation results on our particular dataset.

3.8 Assessing the Reliability of Deep Learning Models

After completing the crucial preprocessing and training, the next step involves evaluating the model's performance on the test set. Performance metrics are crucial for this assessment, and for our case, we adopted metrics [25] such as:

- Precision measures a model's ability to correctly identify positive examples among all examples identified as positive. It is calculated by dividing the number of true positives by the sum of true positives and false positives [25] (see Eq. 7).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

- Recall measures a model's ability to identify all real-world positive examples. It is calculated by dividing the number of correctly identified positive examples by the sum of true positives and false negatives [25] (see Eq. 8).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

- F1-Score is a metric that combines both precision and recall into a single measure using the harmonic mean. It is useful when finding a balance between precision and recall is desired [25].

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

- The Kappa coefficient measures the agreement between the model predictions and the actual classes, taking chance into account. P_o is the observed agreement proportion, and P_e is the expected agreement proportion by chance [32].

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{10}$$

- For the evaluation in each epoch, we use *Accuracy* (see Eq 11) that measures the total proportion of correct predictions among all predictions [1] where TN is the True Negative and the rest is already define upper.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

- Furthermore, we use the $Jaccard$ coefficient (see Eq 12) that measures the similarity between two sets using set intersection and union [24].

$$Jaccard = \frac{TP}{TP + FP + FN} \tag{12}$$

3.9 Process to Detect Changes in Images

After assessing and selecting the best-performing model based on metrics, this section outlines the detailed methodology employed for visualizing temporal changes in images. The approach involves a pixel-wise comparison using specially generated masks from the chosen deep learning model. These masks pinpoint the exact areas of change and quantify the alterations, allowing for a nuanced analysis while retaining semantic information.

3.9.1 Quantification of Changes between Two Dates. As we consider that the area is divided into eight distinct classes: cultivated fields, uncultivated fields, land, water, buildings, roads, football fields, and vegetation. Let $M_{i_{\text{date(i)}}}$ be the set of all classes of an image, i the index of each mask in $date(i)$ such that $date(i)$ a period where image was taken.

Let C be a subset of the mask $M_{i_{\text{date(i)}}}$. For every class c in C and every pixel p in $M_{i_{\text{date(i)}}}$, the percentage at the date i ($date(i)$), noted as $\text{Percent}_{\text{date}(i)}(c)$, is defined as the ratio of the sum of pixels belonging to class C to the sum of all pixels in M.

$$\text{Percent}_{\text{date(i)}}(c) = \frac{\sum\limits_{p \in C} p}{\sum\limits_{p \in M_{i_{\text{date(i)}}}} p} \times 100 \tag{13}$$

Therefore, the process quantifies the pixel distribution for each class in masks captured at different periods, by evaluating the variation of each class following Eq 14. This enables understanding the direction of changes between two dates, indicating whether a class increases or decreases during that period.

$$Diff_{\%}(c) = Percent_{date(2)}(c) - \text{Percent}_{\text{date}(1)}(c) \tag{14}$$

3.9.2 Highlighting Areas that Have Changed. The first step in the process involves creating a changes map using Eq 15. This equation calculates the absolute difference between the number of pixels for each class in two predicted masks, resulting in a map that highlights areas with significant alterations.

$$\text{Mask\_Diff}(M_{2_{\text{date(2)}}}, M_{1_{\text{date(1)}}}) = |M_{2_{\text{date(2)}}} - M_{1_{\text{date(1)}}}| \tag{15}$$

Eq 15 provides a pixel-wise representation of changes, facilitating a detailed understanding of modifications in land cover classes between compared periods. The changes map is then superposed onto the last real image to precisely visualize changed areas.

Table 1: Global metrics

Metrics	Accuracy	Jaccard	Kappa
UNET	96%	92.55%	94.98%
FCN8	91%	84%	88.55%

Table 2: Performances of deep learning models in the confusion matrix.

Models	TP	TN	FP	FN
UNET	96.10%	99.44%	3.89%	0.55%
FCN8	91.15%	98.73%	8.84%	1.26%

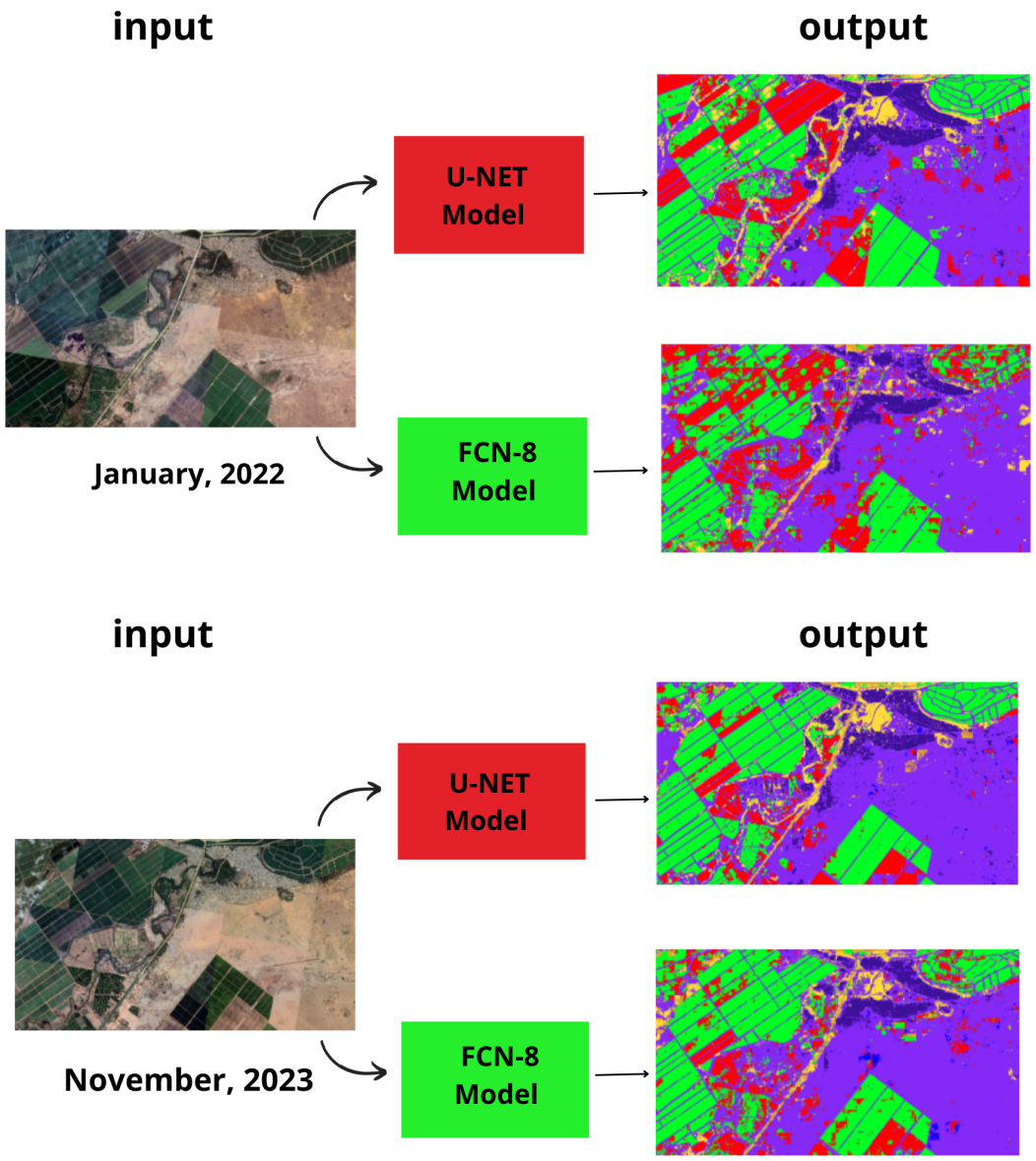


Figure 5: Image segmentation with models in January 2022 and November 2023

4 RESULTS

In this section, we present the results of our study on both models. In Table 3, we have the classification report of models which highlights the metrics such as the precision, the recall and the F1-score that is harmonic mean of both metrics, of each class of our study.

We have organized the global metrics in the Table 1 for both models such as the accuracy that is a measure of the overall correctness of a classification model, the jaccard coefficient that measures the similarity between two sets by calculating the ratio of the size of their intersection to the size of their union. and the kappa coefficient that measures the level of agreement between two raters or, in the context of classification, between predicted and actual labels. It accounts for the possibility of agreement occurring by chance.

In Table 2, we have the performances of deep learning models in the confusion matrix.

We collect two images respectively in january 2022 and november 2023 with 8192x4930 resolution from Google Earth Pro. And we make the segmentation of images by using two models (U-NET and FCN-8) in order to have the segmented images with the different categories within the images for finally make the comparison between both masks to get the difference map. We can see for both models

Table 3: Classification report of models

Metrics	Precision	Recall	F1-score
Classes/Models	U-Net | FCN-8	U-Net | FCN-8	U-Net | FCN-8
Land	0.95 | 0.85	0.93 | 0.88	0.94 | 0.87
Vegetation	0.92 | 0.89	0.96 | 0.87	0.94 | 0.88
Building	0.95 | 0.91	0.97 | 0.91	0.96 | 0.91
Road	0.93 | 0.73	0.97 | 0.78	0.95 | 0.75
Water	0.97 | 0.92	0.95 | 0.86	0.96 | 0.89
Cultivated field	0.98 | 0.97	0.97 | 0.94	0.98 | 0.95
Uncultivated field	0.97 | 0.94	0.99 | 0.94	0.98 | 0.94
Football field	0.96 | 0.96	0.98 | 0.97	0.97 | 0.96

the result of the segmentation over the two images respectively in 2022 and 2023 in Fig 5.

Based on the metrics illustrated in Tables 1, 2 and 3, we note that the U-NET is better for our case so we adopt it to continue the study.

To delve deeper into the analysis, we employ Eq 13 to quantify all classes, obtaining the distribution of each class across different years. This evolution is comprehensively depicted in Table 4, Fig 6, and Fig 7. In Fig 7(a), the distribution is visually presented through a histogram, illustrating the nuanced changes in each class between the years 2022 and 2023.We evaluate the difference in attendance rates for each class between the two respective years (See Eq 14), providing a graphical representation of the rate of changes. In Fig 6, we have the absolute representation of rate of change and we see that the most representative change concerns the uncultivated field with 39.8% of rate of change and less representative is the road with 0.1%. Fig 7(b) provides a visual representation of the rate of change, where negative values for classes such as land, water, cultivated field, and football field indicate a decrease from 2022 to 2023. Conversely, positive values for classes like building, vegetation, and uncultivated field mean an increase during the same period.

We get the segmented images that gives us the semantic of each class, we use the segmented images that are the predicted masks for both different times to create the difference map by doing the difference between them in order to highlight only the changes areas where the white pixels represents the changes and the black pixels represents the non-changes.

After getting the difference map we create the changes map by superposing the white pixels in the difference map with the last image taken like in november 2023. By doing that we achieve to create a changes map where we can see the areas that have changed and are represented by white pixels as we can notice it in the Fig 8

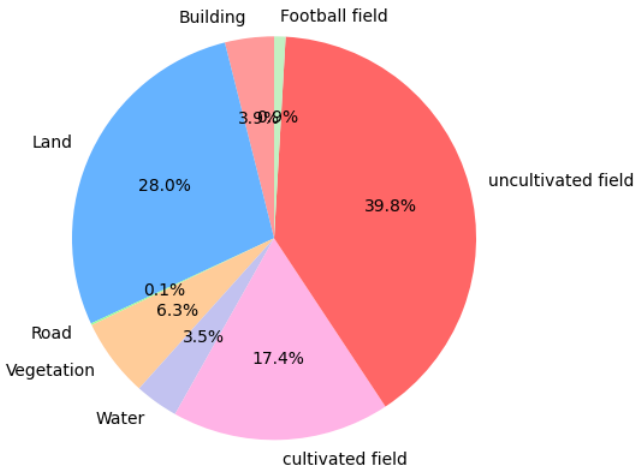
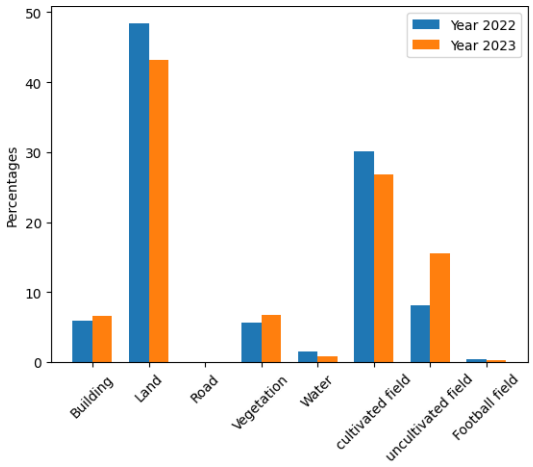


Figure 6: The rate of change in classes between 2022 and 2023

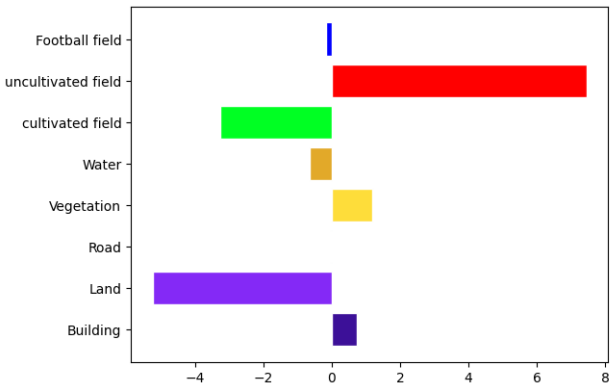
5 DISCUSSION

Our study, centered on evaluating the effectiveness of U-Net and FCN-8 for land cover change detection, stands at the intersection of cutting-edge deep learning techniques and the evolving landscape of remote sensing applications. As we dissect the findings, it's pivotal to contextualize our work within the broader spectrum of existing studies, bringing to light both similarities and distinctions.

The stellar performance of our chosen models, evidenced by high accuracy rates of 96% and 91% for U-Net and FCN-8, respectively, sets a strong foundation for practical applications. The Jaccard and Kappa coefficients further validate the robustness of these models, emphasizing their ability to discern intricate land cover changes with accuracy.



(a) Frequency distributions of classes according to the years



(b) The rate of change between 2022 and 2023 with directions

Figure 7: Classes distribution between 2022 and 2023

Table 4: Frequency distribution table from U-NET masks generated.

Classes/Year	2023	2022
Building	6.6%	5.86%
Land	43.16%	48.4%
Road	0.03%	0.06%
Vegetation	6.74%	5.56%
Water	0.81%	1.46%
Cultivated Field	26.8%	30.06%
Uncultivated Field	15.55%	8.11%
Football Field	0.31%	0.48%

The granular insights provided by the confusion matrix amplify our confidence in the models. With high true positive and true negative rates and minimal false positives and false negatives, U-Net and FCN-8 showcase reliability in classifying positive and negative instances. This bodes well for real-world scenarios, where precision is paramount in domains like urban planning, environmental monitoring, and agricultural management.

Comparing our approach with existing studies reveals a diverse tapestry of methodologies in land cover change detection.

Multimodal Siamese Architecture vs. UNET and FCN8: The first study, leveraging a multimodal Siamese architecture, underscores the adaptability of models to different datasets. In contrast, our choice of U-Net and FCN-8 demonstrates comparable efficacy, revealing the versatility of architectures in capturing complex changes across varied scenarios.

HRMS Images with Pan-Sharpening vs. VHR Multispectral Orthorectified Images: The second study excels with HRMS images and pan-sharpening, whereas the third study emphasizes Very High-Resolution (VHR) multispectral orthorectified images. Our work aligns with the latter, emphasizing the need to validate classifier generalizability on diverse datasets, acknowledging the nuances introduced by varying image qualities.

BiLSTM for Sentinel-2 Time Series Data vs. FCN-8 with VGG-16 Weights: The fourth study's use of BiLSTM for Sentinel-2 time series data and the fifth study's proposition of FCN-8 with VGG-16 weights add richness to the methodological palette. Our alignment with the latter study signals the potential of FCN-8 with VGG-19 in achieving accurate land cover classifications across diverse scenarios.

While U-Net exhibits superior performance compared to FCN-8 in land cover classification, these comparative insights unveil the nuanced strengths of each model within the dynamic landscape of land cover change detection. Our study, therefore, not only contributes valuable performance metrics but also deepens our understanding of the diverse approaches available for tackling this complex challenge.

6 CONCLUSION AND FUTURE WORKS

The study is conducted to detect land use and land cover changes, with a specific focus on mitigating and anticipating social tensions between agricultural and pastoral communities. We delved into the segmentation of the different areas by using two models of deep learning which are U-NET and FCN-8 in order to automate the process and to keep semantic of areas. We notice that the U-NET model is better for our case and is more accurate than FCN-8 and we choose to continue the study with it. After, we make the comparison between two generated masks from U-net model of images taken in different periods to highlight the changes areas and we create changes map by superposing the difference map to the last taken image. We also quantify the rate of changes by using some graphical representations based on the distributions of our different classes in two different dates. The performance of models can be due to how good is the annotation which is an important step for this study. The limitation is the fastidious annotation due the supervised learning and the material resources.

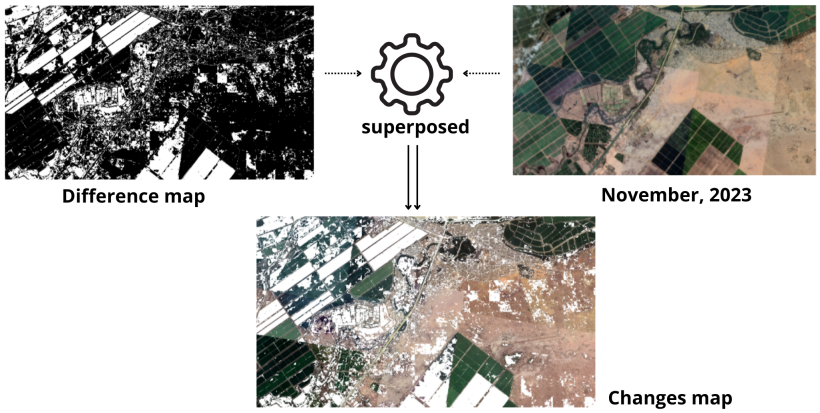


Figure 8: Description of the process for obtaining the change map.

This research can be extended to be align with the objectives of the Great Green Wall (GGW) project [26], aiming to address environmental challenges and promote sustainable land management practices across the region.

As future works, we plan to continue the study by using others images that have more channels to make the model stronger and we can pass from the detection of changes to the prediction of changes by using the distribution data and some techniques of machine learning to know in advance the areas which will probably change in some years.

REFERENCES

[1] Alireza Baratloo, Mostafa Hosseini, Ahmed Negida, and Gehad El Ashal. 2015. Part 1: simple definition and calculation of accuracy, sensitivity and specificity. (2015).

[2] Aniruddha Bhandari. 2023. Feature engineering: scaling, normalization, and standardization (Updated 2023). *Analytics Vidhya*. (2023).

[3] Vitaly Bushaev. 2018. Adam—latest trends in deep learning optimization. *towards data science* 22 (2018).

[4] Abu Bakar Siddik Nayem et al. 2020. LULC Segmentation of RGB Satellite Image Using FCN-8. *CoRR* abs/2008.10736 (2020).

[5] Aryal Jagannath et al. 2023. Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne, Australia. *Scientific Reports* 13, 1 (2023), 13510.

[6] Amit Kumar et al. 2019. Evaluation of Radarsat-2 quad-pol SAR time-series images for monitoring groundwater irrigation. *International Journal of Digital Earth* 12, 10 (2019), 1177–1197.

[7] Chaudhary et al. 2020. Activation functions: SIGMOID, Tanh, Relu, leaky relu, softmax. *Medium* (2020).

[8] Campos-Taberner et al. 2020. Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific reports* 10, 1 (2020), 17188–17198.

[9] Chenchen Wang et al. 2021. A joint change detection method on complex-valued polarimetric synthetic aperture radar images based on feature fusion and similarity learning. *International Journal of Remote Sensing* 42, 13 (2021), 4864–4881.

[10] Destras Oceane et al. 2023. Survey on Activation Functions for Optical Neural Networks. 56, 2 (2023).

[11] Ebel et al. 2021. FUSING MULTI-MODAL DATA FOR SUPERVISED CHANGE DETECTION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B3-2021 (2021), 243–249.

[12] Georganos Stefanos et al. 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE geoscience and remote sensing letters* 15, 4 (2018), 607–611.

[13] Huaqiao Xing et al. 2021. Integrating change magnitude maps of spectrally enhanced multi-features for land cover change detection. *International Journal of Remote Sensing* 42, 11 (2021), 4284–4308.

[14] Kafy Abdulla-Al et al. 2022. Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms. *Building and Environment* 217 (2022), 109066.

[15] Karra Krishna et al. 2021. Global land use/land cover with Sentinel 2 and deep learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS*. IEEE, 4704–4707.

[16] Minaee et al. 2021. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence* 44, 7 (2021), 3523–3542.

[17] Noryusdiana et al. 2017. Towards the use of remote-sensing data for monitoring of abandoned oil palm lands in Malaysia: a semi-automatic approach. *International Journal of Remote Sensing* 38, 2 (2017), 432–449.

[18] Noryusdiana Mohamad Yusoff et al. 2017. Phenology and classification of abandoned agricultural land based on ALOS-1 and 2 PALSAR multi-temporal measurements. *International Journal of Digital Earth* 10, 2 (2017), 155–174.

[19] Russell et al. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision* 77 (2008), 157–173.

[20] Sharma et al. 2021. Identifying Seasonal Groundwater-Irrigated Cropland Using Multi-Source NDVI Time-Series Images. *Remote Sensing* 13, 10 (2021).

[21] Subramanian et al. 2022. On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Computing and Applications* 34, 16 (2022), 13951–13968.

[22] Sawant Suraj et al. 2023. Sen-2 LULC: Land use land cover dataset for deep learning approaches. *Data in Brief* 51 (2023), 109724.

[23] Wang Junye et al. 2022. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Science of the Total Environment* 822 (2022), 153559.

[24] Julian LE GOUIC. 12/07/2022. COMPUTER VISION, MACHINE LEARNING. https://www.quantmetry.com/blog/choix-fonction-de-perte-en-computer-vision-partie-2/.

[25] Rohit Kundu. 2023. F1 score in machine learning: Intro and calculation.

[26] United Nations. Accessed December 6, 2023. Great Green Wall Initiative. https://www.unccd.int/our-work/ggwi.

[27] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. 2017. Activation functions in neural networks. *Towards Data Sci* 6, 12 (2017), 310–316.

[28] Pawel Siciarz and Boyd McCurdy. 2022. U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk in head and neck cancer radiation therapy based on computed tomography scans. *Physics in Medicine and Biology* 67, 11 (jun 2022), 115007.

[29] Qinqin et al. Sun. 2012. The relationship between land surface temperature and land use/land cover in Guangzhou, China. *Environmental Earth Sciences* 65 (2012), 1687–1694.

[30] Nagesh Kumar Uba. 2019. Land Use and Land Cover Classification Using Deep Learning Techniques. arXiv:1905.00510 [eess.IV]

[31] Saad Wazir and Muhammad Moazam Fraz. 2022. HistoSeg: Quick attention with multi-loss function for multi-structure segmentation in digital histology images. In *2022 12th IEEE International Conference on Pattern Recognition Systems (ICPRS)*.

[32] Maarit Widmann. viewed in december, 2023. Cohen's Kappa: What It Is, When to Use It, and How to Avoid Its Pitfalls. https://thenewstack.io/cohens-kappa-what-it-is-when-to-use-it-and-how-to-avoid-its-pitfalls/.

[33] Wahyu Wiratama, Jongseok Lee, and Donggyu Sim. 2020. Change Detection on Multi-Spectral Images Based on Feature-level U-Net. *IEEE Access* 8 (2020), 12279–12289.